International Journal on Advanced Science Engineering Information Technology

Vol.11 (2021) No. 2
ISSN: 2088-5334

Waveforms Classification of Northern Sumatera Earthquakes for New Mini Region Stations Using Support Vector Machine

Marzuki Sinambela[a,c,*], Marhaposan Situmorang[a], Kerista Tarigan[a], Syahrul Humaidi[a], Makmur Sirait[b]

[a]*Department of Phyiscs, FMIPA, Universitas Sumatera Utara, Medan, Indonesia*
[b]*Department of Phyiscs, FMIPA, Universitas Negeri Medan, Medan, Indonesia*
[c]*Badan Meteorologi, Klimatologi dan Geofisika, Indonesia*
*Corresponding author: *marzuki.sinambela@bmkg.go.id*

***Abstract*— We develop and evaluate the new mini region station in Northern Sumatera for discrimination and feature extract seismic events form shallow and intermediate based on waveforms recorded. Machine learning approaches are employed to classification the waveforms and seismic features of the recoded signal in the time-frequency domain. The most issue of this study are the recurrence of the seismic tremors in January to April 2020 regularly happened, and exceptionally local in Northern Sumatra. This can be also in related to the establishment of modern sensors, for that it will be fundamental to create a high-performance technique for automated clustering of seismic tremors recorded of the modern smaller than expected locale sensors on a limited assortment of floor collectors based on their supply depths. We applied the technique to 25 earthquakes that started January to April 2020, with the depth are smaller than 100 km in the land. A selected set of features were then used to train the system to discriminate from events with a hypo-central depth between 10 to 100 km with 96.01 percent accuracy using the SVM model. The result shows that the spectral feature using wavelet-based with machine learning python (mlpy) package has the highest energy correlation. Wavelet spectral in the time-frequency domain is all-new mini region stations that are promising for seismic event classification. The used machine learning approaches have a good classification of low energy signals recorded at the new mini region station in Northern Sumatera.**

***Keywords*—Classification; SVM; wavelet-based; machine learning; new mini region station.**

Manuscript received 11 Jul. 2020; revised 2 Dec. 2020; accepted 24 Feb. 2021. Date of publication 30 Apr. 2021.

I. INTRODUCTION

The new mini region station had been deployed at the end of 2019 and operated now. The new mini region station was deployed to monitor the local earthquake in northern Sumatera. The characteristics of earthquakes in Northern Sumatera usually came from the local earthquake in the land. The earthquake activity is strongly influenced by the Sumatra Fault Zone (SFZ) and local segment [1]. Based on the history of earthquake activity in Northern Sumatra, a global monitoring system and a very local system is needed to detect local earthquake events. The new mini region station at the end of 2019 had been deployed and worked. The new mini region station can detect a very local earthquake in Norther Sumatera. In this study, we used 9 stations located in Northern Sumatra. They are ONSM, PLSM, PTSM, RSSM, SDSM, SPSM, TKSM, and TTSM station. The new mini region installation was operated by the observatory of Indonesia Meteorology, Climatology and Geophysics Agency (BMKG).

The observatories use computerized event detection and area techniques for monitoring nearby and regional seismicity. Fully computerized match solutions supply a cost-effective, almost real-time snapshot on seismic endeavor inside a goal area. However, the events are regularly unclassified or poorly classified. The next and crucial step is to follow reliable automatic or semi-automatic methods for classifying the considerable database of entirely automated event solutions. Automated match classification is quintessential for monitoring natural dangers when rapid and dependable statistics to nearby authorities and media are essential. Moreover, it helps preserve the quality of regional earthquake catalogs, in precise among the low magnitude events. Namely, if unclassified or poorly labeled tournament options quit up in the catalog, the earthquake statistics will turn out to be increasingly contaminated with anthropogenic activity.

For perception, the fundamentals of earthquake waveform networks are very important to classify the waveform primarily based on the fashionable earthquake data. The overall performance of the seismic station had been notably influenced by sign recorded and the information availability. The classification of seismic occasions requires the integration of physical and statistical techniques. The assignment is difficult in low-seismicity areas where herbal and anthropogenic seismicity regularly overlaps in magnitude, space, and time. Sparse insurance of the monitoring network similarly complicates match classification. This study classifies the waveforms from the Northern Sumatra earthquake based on the new mini region station. Classification of waveforms is primarily based on the computer getting to know the method to aid SVM Model and Continuous Wavelet Transform.

SVM model has become one of the most popular machine learning tools in digital signal processing. Discriminating between a range of types of seismic activities is of great scientific, [2]–[4] used a machine learning technique employing SVM to classify the earthquake and seismic data in the region. The automatic classification of the seismic event was studied [5]. Using a supervised pattern recognition technique in SVM, the classification result can perceive and filter blasts and spurious events from thoroughly automatic event options with an excessive accuracy level.

Wavelet-Based computer mastering is an accurate, efficient, and efficacious method to enhance the seismic signal's nice. Classification of the signal from waveforms data depends on considerable lookup in digital sign processing and seismology [6]–[10].

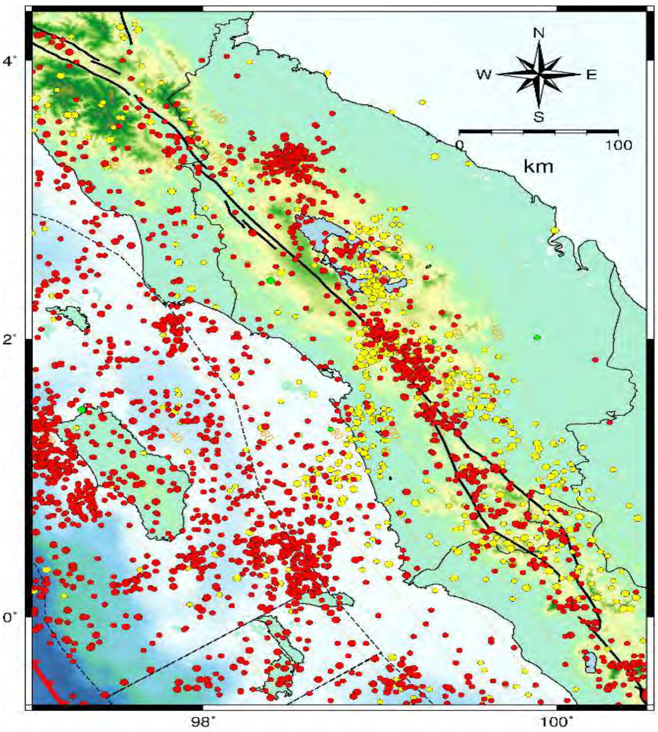


Fig. 1 Seismicity in Northern Sumatera, period 2009 to 2019

The main problem of this study, the frequency of the earthquakes from January to April 2020 often occurred and very local in Northern Sumatra. This is also related to the installation of new sensors, for that it will be necessary to develop a high-performance strategy for automated clustering of earthquakes recorded of the new mini region sensors on a restrained variety of floor receivers based on their supply depths. Based on fig. 1, the characteristics of earthquakes in Northern Sumatra come from shallow (red circle) and intermediate (yellow circle) dominant, with different Magnitude. The earthquake activity was caused by the Sumatera Fault Zone and induced from segment local activity in the Northern Sumatera.

II. Material and Method

For this study, we utilized data from 8 new mini region stations in Northern Sumatera (Fig.2). The selected station was based on the new installation at the cease of 2019 in Northern Sumatera and had an adequate range of earthquake recordings from January to April 2020. In this case, we used supervised learning techniques toward the final solution. The primary is to develop a machine learning model that can distinguish natural earthquakes from northern Sumatra. Three impartial even statistics sets had been compiled for this study. The first set was once used for SVM training, the second for making an attempt out the network processing guidelines, and the third for evaluating the SVM classification average overall performance.

Since there are only two types of data, the model would be the binary classification problem. The performance will be evaluated by accuracy, recall, F-1, and receive operating characteristic (ROC) scores based on the predicted and actual value of the test dataset. The dataset comes from Indonesia Seismic Monitoring's amenities, BMKG-IA, to get entry to the waveforms, associated metadata, and derived products used in this study.

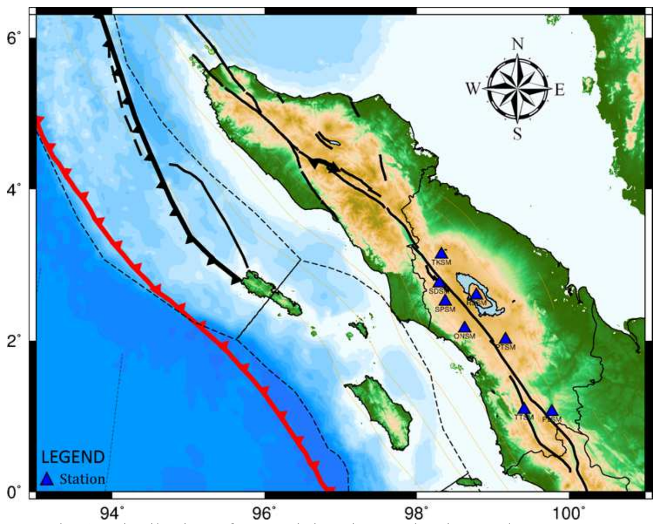


Fig. 2 Distribution of new mini region station in Northern Sumatra

The dataset of the earthquake in this study, compared to shallow and intermediate. Based on the intermediate from January to April 2020, there is seven events in the Northern Sumatera as shown in Table 1. Table 1 indicates the intermediate earthquake from 61 to 100 km. The shallow earthquake from January to April 2020 had been recorded, as shown in Table 2. There are 19 earthquakes recorded between 0 to 60 km.

TABLE I
THE INTERMEDIATE EARTHQUAKE IN NORTHERN SUMATERA, PERIOD JANUARY TO APRIL 2020

Origin Time	Mag.	Type	Lat.	Long.	Depth	Region
2020-03-22T18:03:49	3.6	M	1.24	98.87	76.6	Northern Sumatra, Indonesia
2020-03-16T18:49:22	2.7	M	1.96	99.05	94.8	Northern Sumatra, Indonesia
2020-03-13T23:04:24	3.6	M	1.55	98.82	84.5	Northern Sumatra, Indonesia
2020-02-16T12:59:38	3.1	M	1.32	98.88	83.7	Northern Sumatra, Indonesia
2020-02-10T12:12:38	4.3	M	1.50	98.89	78.9	Northern Sumatra, Indonesia
2020-02-09T17:14:24	3.0	M	1.94	98.91	94.9	Northern Sumatra, Indonesia
2020-01-16T00:14:27	3.2	M	1.12	99.02	94.8	Northern Sumatra, Indonesia

TABLE II
SHALLOW EARTHQUAKE IN NORTHERN SUMATRA, PERIOD JANUARY TO APRIL 2020

Origin Time	Mag.	Type	Lat.	Long.	Depth	Region
2020-05-05T12:31:33	2.6	M	1.01	99.48	10.0	Northern Sumatra, Indonesia
2020-05-03T13:46:19	3.4	M	1.67	99.21	10.0	Northern Sumatra, Indonesia
2020-05-01T22:31:24	2.1	M	1.16	99.50	10.0	Northern Sumatra, Indonesia
2020-04-30T08:20:26	5.4	Mw(mB)	1.17	99.46	10.0	Northern Sumatra, Indonesia
2020-04-30T07:09:00	2.8	M	2.94	98.55	10.0	Northern Sumatra, Indonesia
2020-04-28T07:38:16	3.4	M	2.96	98.55	10.0	Northern Sumatra, Indonesia
2020-04-28T04:47:37	3.2	M	2.92	98.55	10.0	Northern Sumatra, Indonesia
2020-04-21T05:27:10	3.0	M	1.78	99.18	10.0	Northern Sumatra, Indonesia
2020-04-06T17:33:54	3.1	M	1.01	99.55	10.0	Northern Sumatra, Indonesia
2020-04-01T21:42:09	2.2	M	1.75	99.21	10.0	Northern Sumatra, Indonesia
2020-03-31T07:48:07	2.4	M	1.83	99.14	10.0	Northern Sumatra, Indonesia
2020-03-22T19:47:29	2.4	M	2.57	98.32	10.3	Northern Sumatra, Indonesia
2020-03-20T04:56:02	2.7	M	1.81	99.21	10.0	Northern Sumatra, Indonesia
2020-03-10T21:21:40	3.6	M	1.73	99.01	10.0	Northern Sumatra, Indonesia
2020-02-10T18:22:43	3.9	M	3.12	98.66	10.0	Northern Sumatra, Indonesia
2020-02-10T11:24:45	3.9	M	1.89	99.09	10.0	Northern Sumatra, Indonesia
2020-02-09T19:17:55	3.1	M	0.97	99.48	10.0	Northern Sumatra, Indonesia
2020-02-07T03:10:27	3.2	M	2.79	98.41	10.0	Northern Sumatra, Indonesia
2020-02-05T22:21:20	2.4	M	2.15	98.38	48.7	Northern Sumatra, Indonesia

In this study, we used the machine learning technique based on the Support Vector Machine. Machine learning is defined as a mathematical model with several parameters that need to be learned from the data. Cervantes et al. [11] first presented support vector machines and have gained popularity since then. SVM classifiers find a hyperplane that maximizes the separation margin and uses a hinge loss function when the data are not separable. Alternatively, the geometric concept of margin can be viewed as a form of regularization. Previous work has shown the equivalence between support vector machines and a robust formulation of the hinge loss classifier [12]. In this paper, we develop new robust formulations for SVM and other classifiers, which lead to further gains in out-of-sample accuracy compared to non-robust methods. The Support Vector Machine (SVM) Model has some hyper-parameters and finding optimal hyper-parameter using the Grid-Search. Furthermore, using the robust optimization for classifying the dataset.

Generating adversarial examples of computing devices gaining knowledge of strategies is essential for designing an extra-strong classifier. Robust assist vector computing device for classification and computational provides decision characteristic immune to statistics perturbations [14]. In this section, the waveforms data the place coaching information are either linearly separable or non-linearly separable, respectively, and furnish computational consequences for real records units.

This paper used SVM to offer a perfect fit for each parameter combination for classifying all sensors for recorded the earthquake from January to April 2020. We still use the approach as a version of stochastic neighbor embedding (t-SNE) [15] to visualize the high-dimensional waveforms data.

Other methods for extracting and analyzing the waveforms data using wavelet-based machine learning to computed the waveforms in machine learning python (Mlpy) in time-frequency [16]. Mlpy is the module for desktop learning constructed on the pinnacle of NumPy/Scipy and the GNU, and we mix the mlpy to extract the nuclear waveform and plotting the spectrum based on continuous wavelet seriously change [17].

A regularly used complicated wavelet in CWT analysis is the Morlet wavelet, which is used during this study as its periodic sinusoidal form capacity that wavelet scales can very fairly be approximated in terms of Hz frequency and interpretation can, therefore, be linked to the existing giant physique of work in seismology based on Fourier frequencies [18], [19].

III. RESULTS AND DISCUSSION

The result of this study, shown in Fig 3, based totally on the waveforms facts recorded, the new mini vicinity stations detected the shallow (red circle) and intermediate (yellow circle) earthquake from January to April 2020, in Fig. 3 indicated the distribution of earthquake typically held inland, with very neighborhood earthquake. The results and discussion may be presented separately or in one combined section and may optionally be divided into headed subsections.

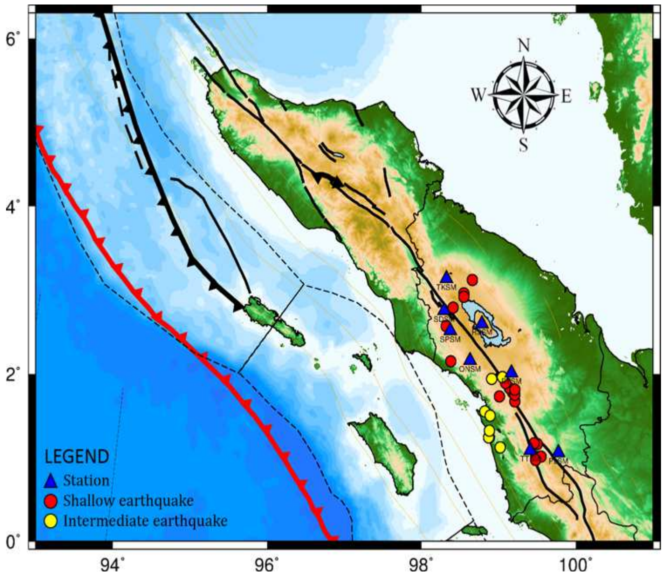
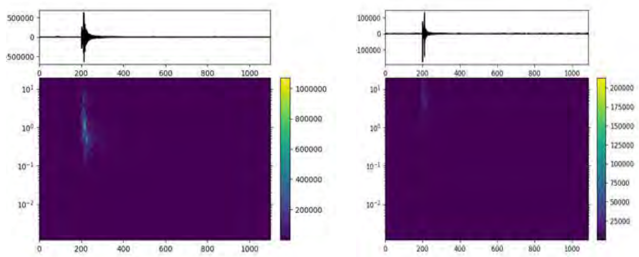
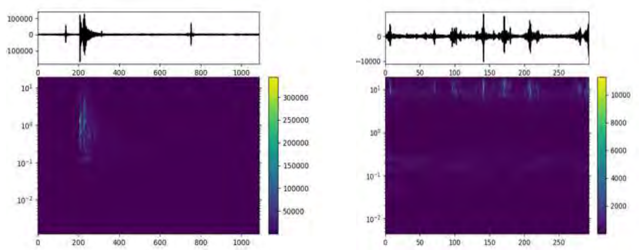


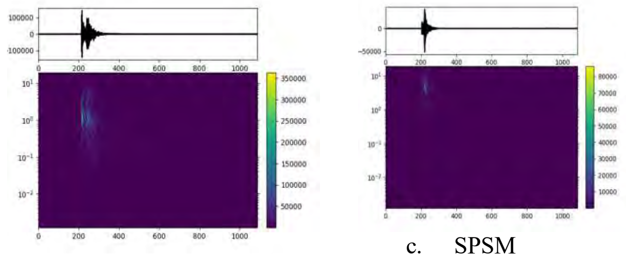
Fig.3. The Shallow and Intermediate earthquake with the new mini region station recorded.

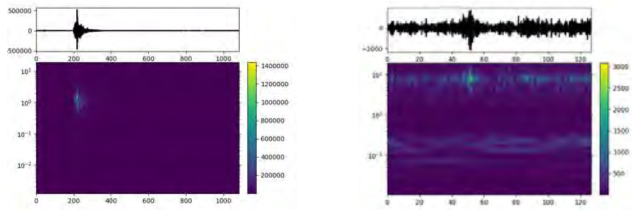
The classification waveforms in the machine are getting to know research, models that predict an outcome from a set of categories shallow and intermediate earthquakes are often referred to as classifiers and projects of predicting an outcome from earthquakes. The seismogram's feature extract shows the total of the intermediate earthquake detected 506 channel and for the shallow is 186 channels with the event earthquake in Northern Sumatera from January to April 2020.

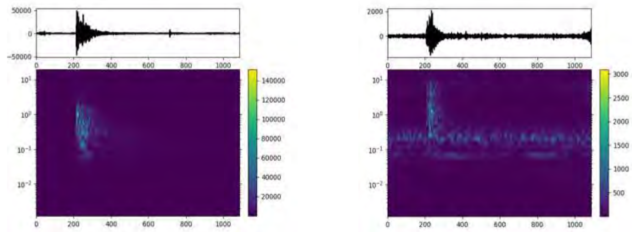
In this study, the larger the feature space, the extra possible mixtures of facets want to be examined and learn waveforms with the machine learning algorithm's aid. There are many unique attributes of the waveforms that have been extracted from the facts in time, frequency, and time-frequency domains. We then check which elements fantastically signify waveforms' characteristics and restrict the classification step to these features.

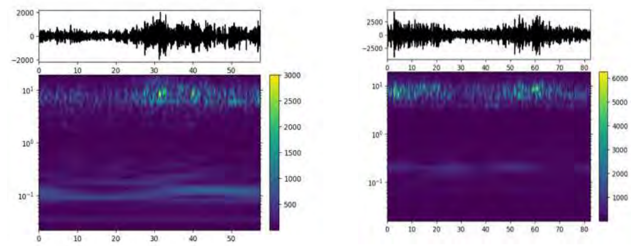

a. LTSM a. LTSM

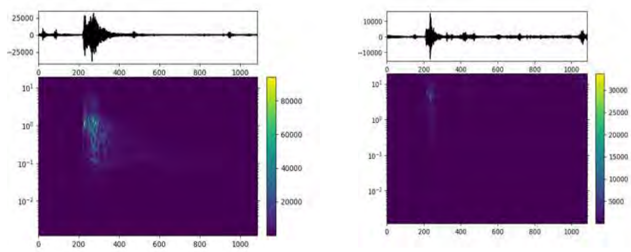

b. ONSM b. ONSM


c. SPSM c. SPSM


d. PTSM d. PTSM


e. RSSM e.RSSM


f. SDSM f. SDSM


g. TKSM g. TKSM

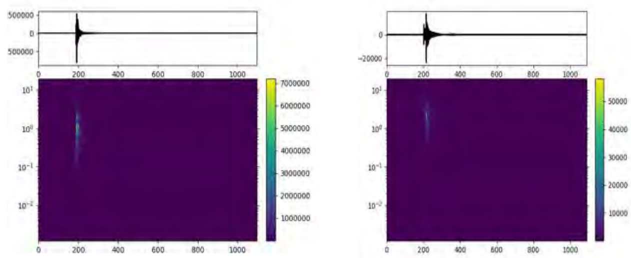

h. TTSM h. TTSM

Fig. 4. The Wavelet-Based plot for the new mini region stations analysis: left and right columns show shallow and intermediate earthquake, respectively.

Figure 4 indicates a selected set of seismograms and their continuous wavelet transform (CWT) spectrograms based on the new mini region stations. Both intermediate and shallow earthquakes are characterized by using floor waves with dominant strength targeted rounds scale 200 and longer periods. The performance of the continuous wavelet transforms (CWT) for spectral representations of shallow and intermediate-seismic signals. Examination of spectrum for the common local earthquake in Northern Sumatera exhibit that the CWT scalograms have better Time-Frequency decision throughout broader frequency tiers than Fourier seriously change spectrograms, which go through from increased spectral smearing in the time domain at greater frequencies.

In this study, the new mini region stations recorded had been computed in spectral. We conclude that the wavelet seriously changes underutilized in a nearby earthquake in Northern Sumatera, where their time-frequency localization houses would be specifically well-suited and potential in terms of automated earthquake detection classification. Intermediate and shallow earthquakes radiate relative greater frequency power for all waveforms of the new mini region stations (Fig.4 a-h).

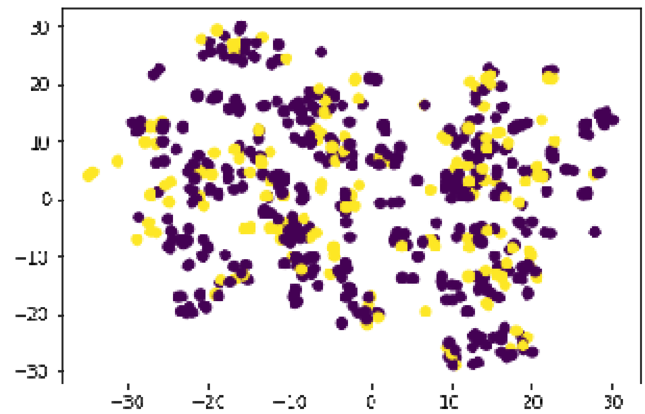


Fig. 5. t-SNE map

The projected data of waveforms for a new mini region station is shown in Fig. 5 as plotting to t-SNE. The visualization is scattered plots representing dimensional rankings of projections of high dimensional characteristic vectors onto the first (x-axis) in opposition to the 2d (y-axis) estimated component.

TABLE III
PERFORMANCE OF THE SVM MODEL

Model	Precision	Recall	F1-Score	Support
Shallow	0.95	0.99	0.97	506
Intermediate	0.97	0.86	0.91	186
Micro avg	0.96	0.96	0.96	692
Macro Avg	0.96	0.93	0.94	692
Weighted	0.96	0.96	0.95	692

The performance evaluation of the SVM model in this study shows in Table 1. After we calculated the percentage of correctly classified instances, we got an accuracy of 95.52 %. In the shallow earthquake, the fraction of accurate (Precision) prediction is 0.95, the fraction of situations that are precisely estimated (Recall) is 0.99, and F measure of test accuracy 0.97, with 506 waveforms recorded. In table 1, exhibit that the intermediate earthquake of the precision is 0.97, Recall is 0.86, and F-measure is 0.91.

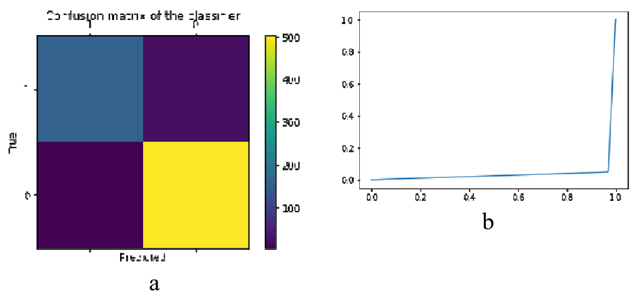


Fig. 6. Confusion Matrix and ROC score

Figure 6 is the confusion matrix from the consequences of the SVM classification suggests the True-P (TP) value of 160 contributions in blue, True-Negative (TN) of 501 in yellow, False Negative (FN) of 5, and False-Positive (FP) of value 26. The matrix confusion in Figure 6 shows the true level of 692 cases. This training data shows the training data is much better for events related to earthquakes on land and at sea. The receive operating characteristic (ROC) is 96.01%, as shown in Figure 6 b.

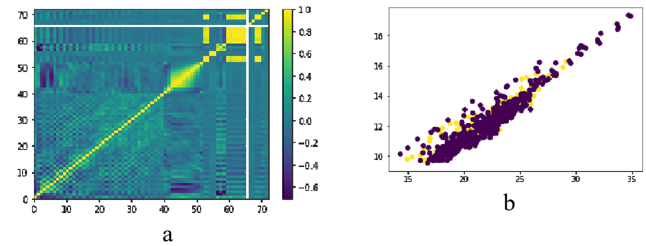


Fig. 7. Visual correlation the feature; a. color bar of t-SNE, b. visualization by isomap

It visualizes high-dimensional waveforms records with the aid of the capability of giving every truth factor a place in a two-dimensional map. Fig. 7a is the correlation among the elements in the shade bar of the approach is a version of stochastic neighbor embedding (t-SNE). The overall performance of t-SNE on a wide range of waveforms statistics in Figure 7 is higher existing strategies at developing a single map that exhibits structure at many one-of-a-kind scales. Figure 7b suggests that the solid universal performance of t-SNE in contrast to Isomap in described founded on modeling large distances of waveforms data. The experiments on various waveforms records devices show off that t-SNE out-performs existing kingdom of the artwork strategies for visualizing various waveforms facts units.

IV. CONCLUSIONS

We use the support vector machine (SVM) approach and Wavelet-based to classify the local earthquake in Northern Sumatera, based on the intermediately and shallow earthquakes. Each of the SVM and Wavelet Classifiers is derived based on the discriminant aspects of the waveform's records in a new mini area station; with the discriminant characteristic incorporates a non-compulsory parameter and spectral amplitude. Primarily based on the chosen set capabilities changed into then used to teach the device to discriminated from sports with a hypo-crucial depth between

10 to 100 km with 96.01 percent accuracy the usage of the SVM model respectively and the wavelet spectral in the time-frequency location is all-new mini region stations that showed to be good measures for seismic occasion class. The computing device mastering techniques have an ideal classification of low power indicators recorded at the new mini region station in Northern Sumatera.

ACKNOWLEDGMENT

The author is grateful to Indonesia Meteorology, Climatology and Geophysics Agency and Geophysics Station of Deli Serdang, Indonesia, to provide and deploy the new mini regions stations and data collection.

REFERENCES

[1] D. H. Natawidjaja, "Updating active fault maps and sliprates along the Sumatran Fault Zone, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 118, no. 1, 2018, doi: 10.1088/1755-1315/118/1/012001.

[2] A. Galkina and N. Grafeeva, "Machine learning methods for earthquake prediction: A survey," *CEUR Workshop Proc.*, vol. 2372, pp. 25–32, 2019.

[3] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Earthquake prediction model using support vector regressor and hybrid neural networks," *PLoS One*, vol. 13, no. 7, pp. 1–22, 2018, doi: 10.1371/journal.pone.0199004.

[4] W. Li, N. Narvekar, N. Nakshatra, N. Raut, B. Sirkeci, and J. Gao, "Seismic data classification using machine learning," *Proc. - IEEE 4th Int. Conf. Big Data Comput. Serv. Appl. BigDataService 2018*, pp. 56–63, 2018, doi: 10.1109/BigDataService.2018.00017.

[5] J. Kortström, M. Uski, and T. Tiira, "Automatic classification of seismic events within a regional seismograph network," *Comput. Geosci.*, vol. 87, pp. 22–30, 2016, doi: 10.1016/j.cageo.2015.11.006.

[6] K. Tarigan, M. Sinambela, M. Panjaitan, P. Simangunsong, and H. K. Siburian, "Machine Learning for Waveform Spectral Analysis on Nuclear Explosion Signal and Performance of Broadband Vertical Component," *J. Phys. Conf. Ser.*, vol. 1120, no. 1, pp. 4–9, 2018, doi: 10.1088/1742-6596/1120/1/012083.

[7] E. Darnila, M. Ula, K. Tarigan, T. Limbong, and M. Sinambela, "Waveform analysis of broadband seismic station using machine learning Python based on Morlet wavelet," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, 2018, doi: 10.1088/1757-899X/420/1/012048.

[8] K. Tarigan, M. Sinambela, M. Panjaitan, P. Simangunsong, and H. K. Siburian, "Machine Learning for Waveform Spectral Analysis on Nuclear Explosion Signal and Performance of Broadband Vertical Component," *J. Phys. Conf. Ser.*, vol. 1120, no. 1, pp. 1–6, 2018, doi: 10.1088/1742-6596/1120/1/012083.

[9] E. Darnila, M. Ula, K. Tarigan, T. Limbong, and M. Sinambela, "Waveform analysis of broadband seismic station using machine learning Python based on Morlet wavelet," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, 2018, doi: 10.1088/1757-899X/420/1/012048.

[10] M. Sinambela, K. Tarigan, and S. Humaidi, "Classification of Seismic Signal by Evaluating Broadband Networks Station in Sumatera Fore-Arc Classification of Seismic Signal by Evaluating Broadband Networks Station in Sumatera Fore-Arc," 2020, doi: 10.1088/1742-6596/1485/1/012054.

[11] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, no. xxxx, 2020, doi: 10.1016/j.neucom.2019.10.118.

[12] M. Singla, D. Ghosh, and K. K. Shukla, "A survey of robust optimization-based machine learning with special reference to support vector machines," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 7, pp. 1359–1385, 2020, doi: 10.1007/s13042-019-01044-y.

[13] H. Xu, "Robustness and Regularization of Support Vector Machines," *J. ofMachine Learn. Res. 10*, vol. 10, pp. 1485–1510, 2009.

[14] F. Ranzato and M. Zanella, "Robustness Verification of Support Vector Machines," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11822 LNCS, pp. 271–295, 2019, doi: 10.1007/978-3-030-32304-2_14.

[15] J. D. Curuksu, *Principles of Data Science: Advanced*. 2018.

[16] M. Sinambela, K. Tarigan, M. Situmorang, S. Humaidi, T. Rahayu, and A. R. Ainun, "Time-Frequency Analysis of Digital Broadband Station Based on Padang Lawas Earthquake," *Mecn. 2020 - Int. Conf. Mech. Electron. Comput. Ind. Technol.*, pp. 69–73, 2020, doi: 10.1109/MECnIT48290.2020.9166633.

[17] M. Sinambela, M. Situmorang, K. Tarigan, S. Humaidi, and K. Sebayang, "Classification of Nuclear Explosion Earthquake Signal based on Machine Learning Approach," *2019 Int. Conf. Comput. Sci. Inf. Technol. ICoSNIKOM 2019*, pp. 2017–2020, 2019, doi: 10.1109/ICoSNIKOM48755.2019.9111479.

[18] M. Sinambela, K. Tarigan, S. Humaidi, and M. Situmorang, "Wavelet based machine learning approach for spectral seismic signal analysis: A case study North Tapanuli earthquake," *AIP Conf. Proc.*, vol. 2221, no. March, 2020, doi: 10.1063/5.0003129.

[19] H. Sirait *et al.*, "Time Frequency Signal Classification Using Continuous Wavelet Transformation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 851, no. 1, 2020, doi: 10.1088/1757-899X/851/1/012045.